hellö

kaan kudret kölköy
devops engineer @ countly

python'da multi-threading ile web crawling

web crawling

thread

python'da web crawling

```python
import urllib.request
import time
import shutil

def downloadImage(filename):
    currentLink = baseURL + filename + '.png'
    currentFile = 'icons/' + filename + '.png'

    req = urllib.request.Request(currentLink, headers = {'User-Agent': 'Mozilla/5.0'})
    with urllib.request.urlopen(req) as res, open(currentFile, 'wb') as outFile:
        shutil.copyfileobj(res, outFile)

baseURL = 'https://dl.kaankolkoy.net/presentation/hacknbreak/lightning-python-threading/icons/'
start   = time.time()

for i in range(1, 21):
    downloadImage(str(i))

print('time elapsed: ' + str(time.time() - start))
```

single thread

```
[lightningtalk-python-threading] python single.py
time elapsed: 4.152761459350586
```

python'da web crawling

```python
import urllib.request
import time
import threading
import shutil


def downloadImage(filename):
    currentLink = baseURL + filename + '.png'
    currentFile = 'icons/' + filename + '.png'

    req = urllib.request.Request(currentLink, headers={'User-Agent': 'Mozilla/5.0'})
    with urllib.request.urlopen(req) as res, open(currentFile, 'wb') as outFile:
        shutil.copyfileobj(res, outFile)

baseURL = 'https://dl.kaankolkoy.net/presentation/hacknbreak/lightning-python-threading/icons/'
start    = time.time()

for i in range(1, 21):
    threading.Thread(target = downloadImage, args = (str(i),)).start()

print('time elapsed: ' + str(time.time() - start))
```

multi thread

```
[lightningtalk-python-threading] python multi.py
time elapsed: 0.10421490669250488
```

sonuç

```
[lightningtalk-python-threading] bash test.sh
# single thread #
time elapsed: 4.481782913208008
file count: 20

# multi thread #
time elapsed: 0.10880541801452637
file count: 20
```

teşekkürler

@kaanklky